Visualization of Valid Polymorphisms within the Human Genome

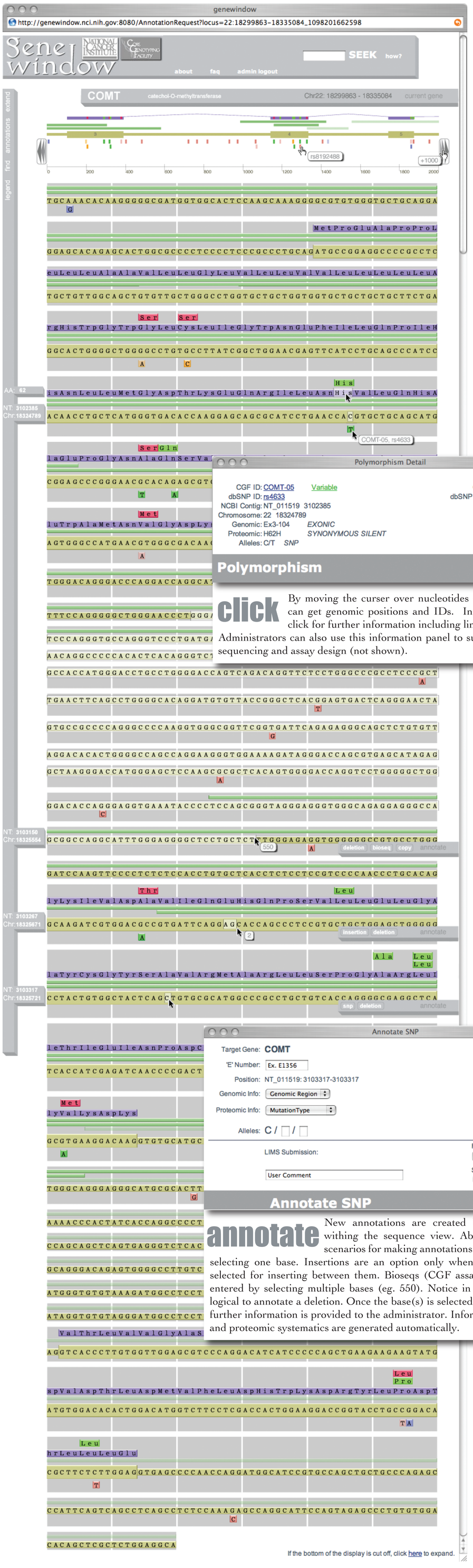
Brian Staats, Liqun Qi, Michael Beerman, Hugues Sicotte, Laura A. Burdett, Bernice Packer, Stephen J. Chanock[1], Meredith Yeager

Intramural Research Support Program, SAIC-Frederick, NCI-FCRDC, Frederick, MD

[1]Section on Genomic Variation, Pediatric Oncology Branch, National Cancer Institute, National Institutes of Health, Gaithersburg, MD

abstract

There are many web-based genome browsers available for researchers interested in human polymorphisms, each with their own unique graphical interface and source for the annotation data. It is typically difficult for researchers, and the public at large, to use these browsers due to the complexity of the data and the limitations of current visualization techniques. A new genome browser has been developed by the Core Genotyping Facility (CGF) at the National Cancer Institute for viewing public annotations together with polymorphisms validated by the CGF. Users can graphically view human variation in genomic and proteomic contexts along with population-specific information, enabling them to make informative decisions based on the polymorphisms that are present in a given population. The database for this browser maintains annotations from many public sources such as NCBI and combines validated internal information on existing and new variations from the CGF. This browser is an integral part of the CGF's pipeline for automating assay design and validation of variations related to cancer studies. Ongoing development includes visual analysis from genotype data and additional information from other public resources.

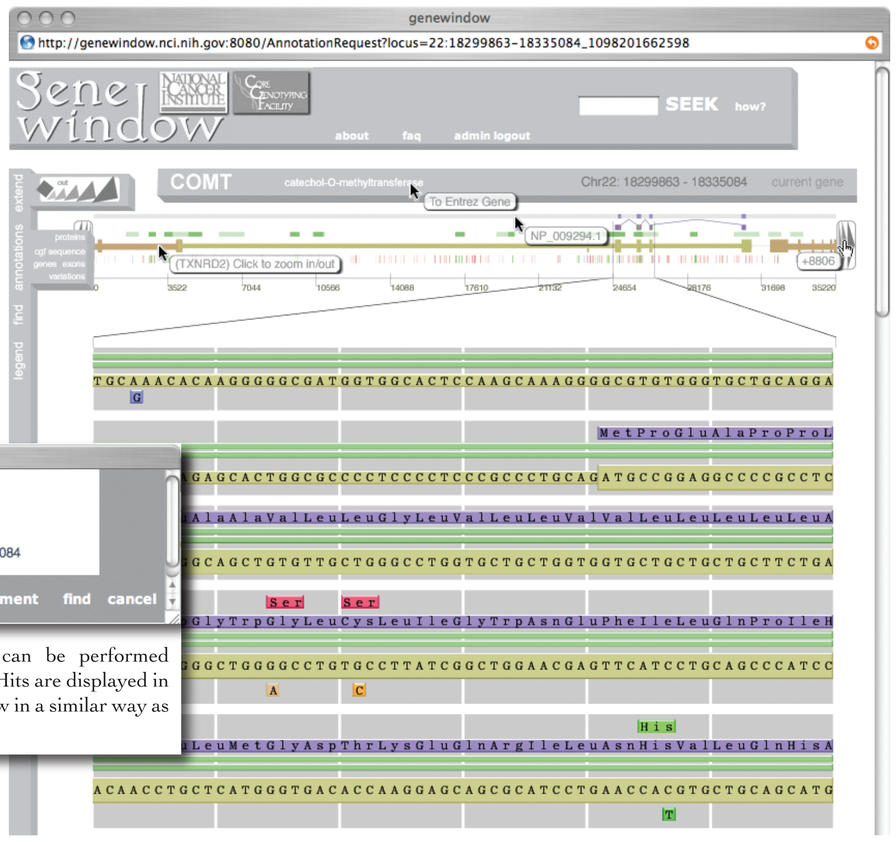


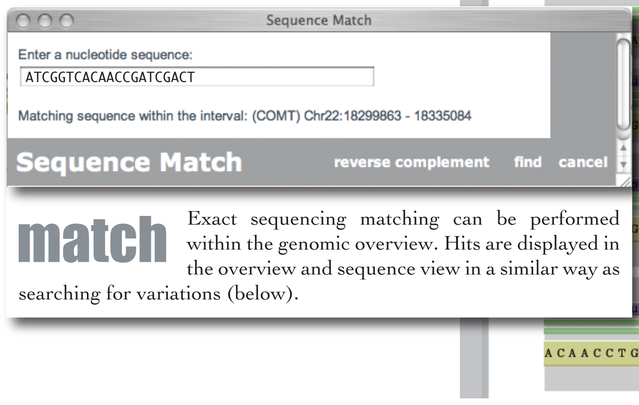
click

By moving the curser over nucleotides or variations, users can get genomic positions and IDs. In addition users can click for further information including links and frequencies. Administrators can also use this information panel to submit variations for sequencing and assay design (not shown).

views

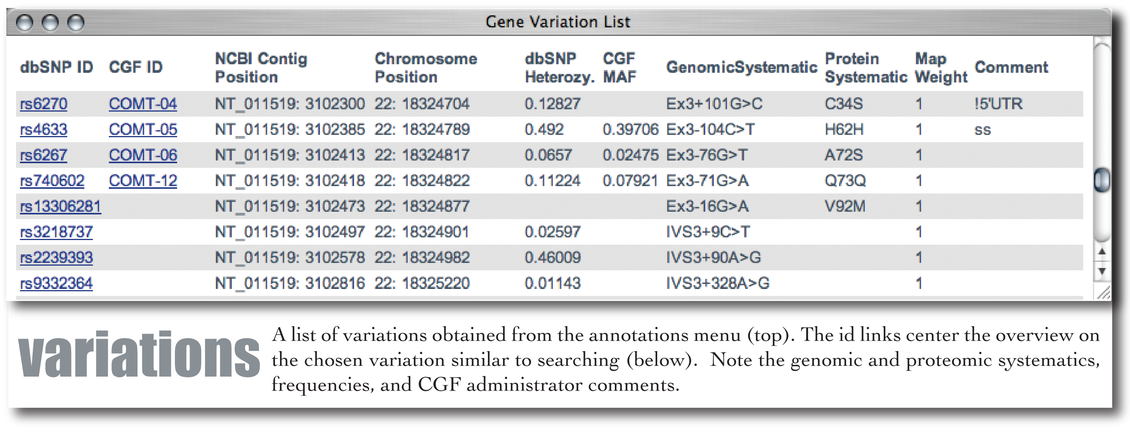
Genewindow is split into two views, an overview that varies in size depending on the gene or genomic region being viewed and, below it, a sequence view displaying a 2000bp region within the overview. Switching the sequence view to display different regions within the overview is accomplished by clicking along the gene in the overview to zoom in and out. The overview displays alternate translations (light purple), current protein (purple), regions sequenced by the CGF (green), current gene (yellow), neighboring genes (brown), and polymorphisms color-coded based on CGF's assay status. The sequence view is similar except only the current protein is displayed. The vertical and diagonal lines between the two views indicate their spatial relationship. Both views illustrate contextual information that appears based on mouseing over objects. The far left bar is the menu where users can expand the genomic interval, access lists of features, find a sequence segment within the overview (match), or view the legend.





match

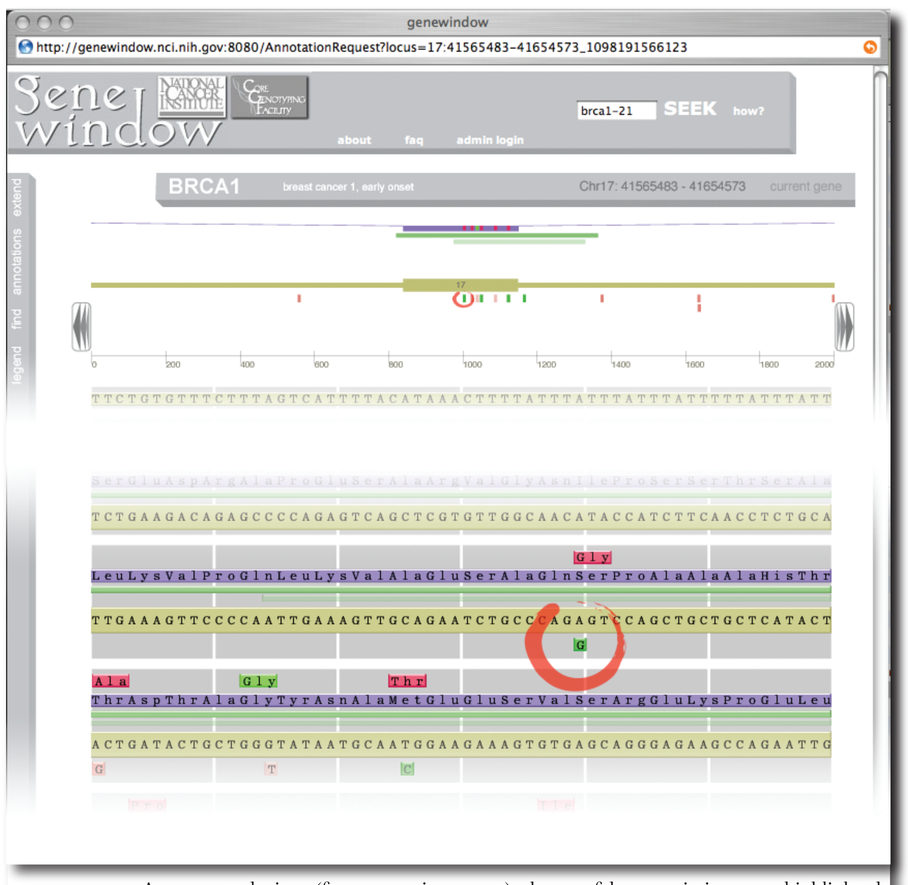
Exact sequencing matching can be performed within the genomic overview. Hits are displayed in the overview and sequence view in a similar way as searching for variations (below).



dbSNP ID	CGF ID	NCBI Contig Position	Chromosome Position	dbSNP Heterozy.	CGF MAF	GenomicSystematic	Protein Systematic	Map Weight	Comment
rs6270	COMT-04	NT_011519: 3102300	22: 18324704	0.12827		Ex3+101G>C	C34S	1	!5'UTR
rs4633	COMT-05	NT_011519: 3102385	22: 18324789	0.492	0.39706	Ex3-104C>T	H62H	1	ss
rs6267	COMT-06	NT_011519: 3102413	22: 18324817	0.0657	0.02475	Ex3-76G>T	A72S	1	
rs740602	COMT-12	NT_011519: 3102418	22: 18324822	0.11224	0.07921	Ex3-71G>A	Q73Q	1	
rs13306281		NT_011519: 3102473	22: 18324877			Ex3-16G>A	V92M	1	
rs3218737		NT_011519: 3102497	22: 18324901	0.02597		IVS3+9C>T		1	
rs2239393		NT_011519: 3102578	22: 18324982	0.46009		IVS3+90A>G		1	
rs9332364		NT_011519: 3102816	22: 18325220	0.01143		IVS3+328A>G		1	

variations

A list of variations obtained from the annotations menu (top). The id links center the overview on the chosen variation similar to searching (below). Note the genomic and proteomic systematics, frequencies, and CGF administrator comments.



seek

A segmented view (for conserving space) shows of how variations are highlighted (circled) as a result of a user's query. Users can search by gene HUGO symbol, dbSNP id, CGF id, or by ranges such as a chromosome and NCBI NT contig intervals. Notice the overview is zoomed in by default and centered around the targeted variation.

conclusion

The mission of the NCI-CGF is to provide quality high-throughput candidate gene SNP genotyping to NCI investigators for molecular epidemiological studies in cancer. Genewindow was developed in response to a need for a tool for use within the complicated workflow of the NCI-CGF. Faced with the need for a central mechanism to view and catalog thousands of SNPs within hundreds of genes, and to track all of the critical associated annotations, we developed Genewindow as an in-house solution. Within our laboratory, it is used not only as a tool for viewing SNPs at genes of interest, it has also been implemented to communicate with our Laboratory Information Management System (LIMS). Recognizing its usefulness and potential for application and/or implementation in other laboratories, we announce it as a publicly available resource (http://genewindow.nci.nih.gov/). Genewindow is a unique tool that may be used by outside investigators in two key ways. First, it is an easy-to-navigate web-based tool that displays genomic annotations in a gene-centric fashion that may enhance decision making for the study of variations at candidate genes. Though there are many other free tools available, Genewindow's simple and easy-to-navigate interface is attractive to geneticists as well as non-geneticist users. Secondly, Genewindow will also be soon available to the open source community for implementation in other government and academic laboratories.

genewindow.nci.nih.gov